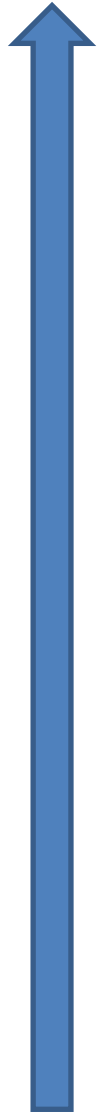


A platform for assessing the current state of data repositories in Korea, Japan and China

Open Repositories 2014
June 9-13, Helsinki, Finland

Suntae Kim
stkim@kisti.re.kr



To analyze the data & identify 6 kinds of features

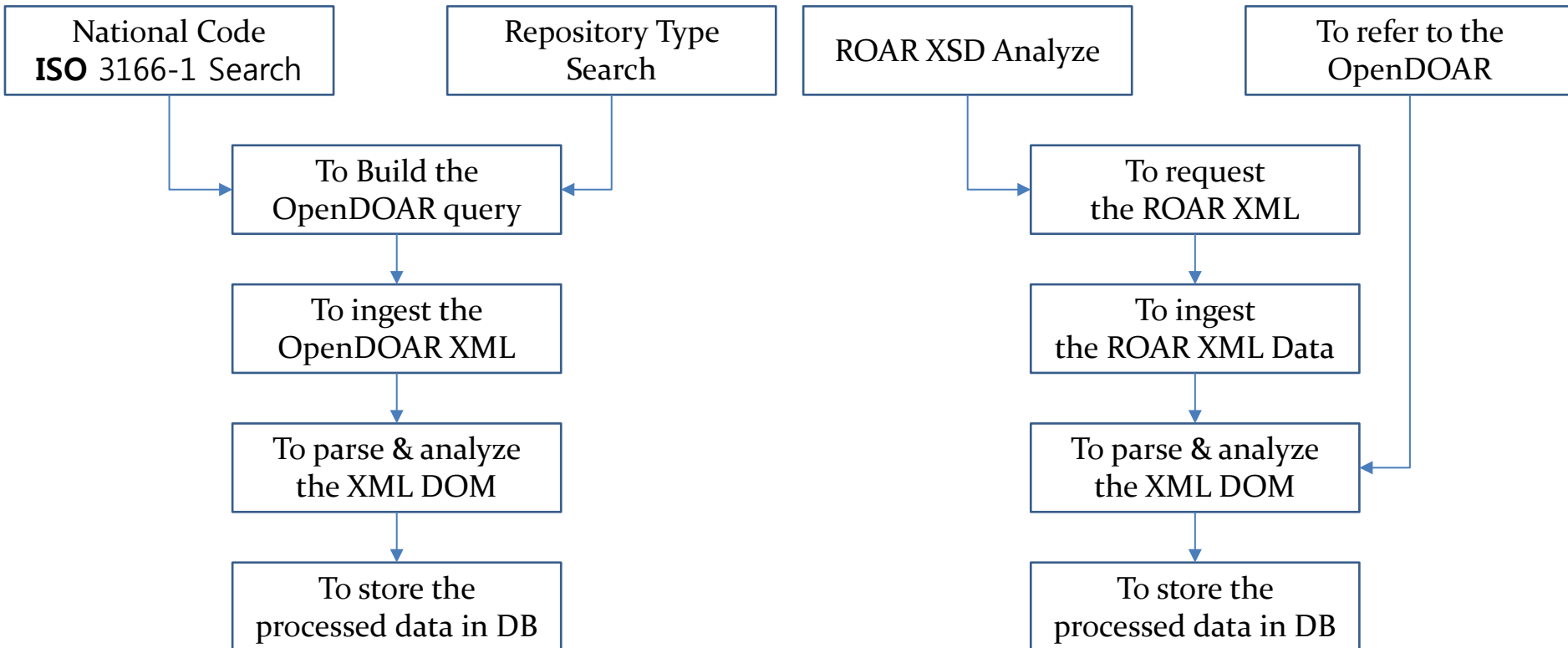
- China, Japan and Republic of Korea operate 288 repositories (42.2% of Asia, 8% of global)
- The repositories that provide the content in Korean was 0.72%.
- The 2,519 repositories are registered in OpenDOAR and ROAR registry services.
- Korea has incomplete repository operation in subject area and paper.
- Korea and Japan, the repositories in 'Health and Medicine' area are in the top.

To identify the collected data

To collect the repository information

To develop the OpenDOAR & ROAR Crawlers

Work flow of the Crawlers

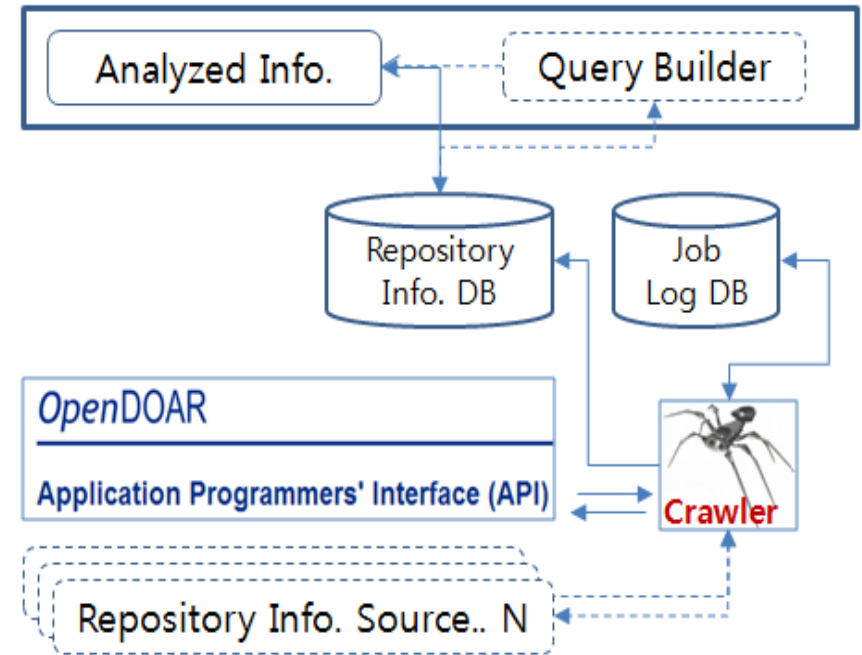


Work flow of the OpenDOAR Crawler

Work flow of the ROAR Crawler

Development Environment

- **Programming language:** Java
- **Java virtual machine:** jdk1.7.0_45
- **DBMS:** MySQL 5.6.15
- **Execution environment:** OS - windows 7
- **Editing environment:** eclipse (Version: Indigo Service Release 2 / Build id: 20120216-1857)



G_itemCnt	K_itemCnt	K_percent	C_itemCnt	C_percent	J_itemCnt	J_percent
164400861	1332674	0.81	472803	0.29	1416966	0.86

language	Repository Cnt	Nation	Repository Cnt
Chinese	106	China	37
Japanese	144	Japan	143
Korean	17	Korea, Republic of	14

global_Rcnt	asia_Rcnt	asia_percent	cjk_Rcnt	cjk_percent_inGlobal	cjk_percent_inAsia
2514	446	17.74	194	7.72	43.50

Repository_Type	Global_Rcnt	Asia_Rcnt	Korea_Rcnt	China_Rcnt	Japan_Rcnt
Aggregating	93	5	(NULL)	(NULL)	5
Disciplinary	275	4	3	(NULL)	1
Governmental	72	1	1	(NULL)	(NULL)
Institutional	2074	184	10	37	137

G_subjectCnt	K_subjectCnt	K_percent	C_subjectCnt	C_percent	J_subjectCnt	J_percent
29	13	44.83	26	89.66	22	75.86

The status of data repositories in each country can be analysed in real time and topics that require more support from data repositories can be identified using the established platform.

Using **Rest-full API from OpenDOAR**, structured information on global repositories is collected and **it is utilized to analyse the status of Korean, Chinese, and Japanese repositories** in comparison with global and Asian ones.

For future study, a module for collecting repository information from other than OpenDOAR needs to be developed.

In addition, development of a Query Builder module and user-interface that can enable multi-dimensional analysis should be followed.